2

NASA TECHNICAL TRANSLATION NASA TT F- 14484

COMBINATORIAL CODING FOR DISCRETE STATIONARY SOURCES

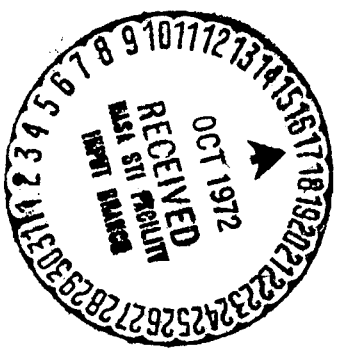
by
Yu. M. Shtar'kov, V. G. Babkin

(NASA-TT-F-14484) COMBINATORIAL CODING FOR DISCRETE STATIONARY SOURCES Y.M. Shtarkov, et al (Agnew Tech-Tran, Woodland Hills, Calif.) Mar. 1972 30 p CSCL 09D G3/08

N72-32209

Unclas 43329

Translation of "Kombinatornoye kodirovaniye dlya diskretnykh statsionarnykh istochnikov", Moscow, Institute of Space Research, Academy of Sciences USSR, Report PR-67, 1971.



Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546 MARCH 1972

30 P.

COMBINATORIAL CODING FOR DISCRETE STATIONARY SOURCES

Kombinatornoe kodirovanie dlya diskretnykh statsionarnykh istochnikov

Yu. M. Shtar'kov, V. G. Babkin

Institute of Space Research of the USSR Academy of Sciences

PR-67

Translated by AGNEW TECH-TRAN
5115 Serrania Avenue
Woodland Hills, California 91364

Prepared for and issued by the Jet Propulsion Laboratory
California Institute of Technology, Pasadena, California
Under NASA contract NAS 7-100

COMBINATORIAL CODING FOR DISCRETE STATIONARY SOURCES

Yu. M. Shtar'kov, V. G. Babkin

A method of nonuniform coding is suggested. The set of messages is divided into non-intersecting classes. Prefix codes are used to describe the classes and the messages within each class. The complete code combination of a message is formed by a successive recording of code combinations of a class and the message within the class.

The expression for the mean length of the code word n using this type of coding of the set of simple messages makes it possible to formulate the specifications for the initial codes. Analogous results have also been obtained for the sequence of messages of length N at the output of a random discrete stationary source. It was shown that for a Markov source of finite connectedness with an unknown matrix of transitions, it is possible to implement a division into classes so that the uniform coding of classes and messages within the class ensures a convergence of the mean number of symbols over message n/N to the entropy of the source, with the growth of N according to the law c log N/N, where c is a constant, which depends on the number of states and the connectedness of Markov's chain. For a random discrete stationary source with unknown statistics, a coding method is suggested in which $\bar{n}/N$ tends to the entropy of this source with the increase of N.

******

Fitingof [1,2] was the first to use combinatorial methods for proving the possibility of efficient coding of the source of independent messages with unknown probabilities of their occurence. Article [3] presents a concrete method of coding for the same case, based on combinatorial considerations and satisfying the asymptotic lower limit of Krichevskij [4].

The suggested method combines combinatorial and probabilistic concepts and consists of grouping the messages at the output of the source, which allows a simple utilization of the available a priori information. Here an especially important role is played by the knowledge of the source model, which determines the interaction of the probabilities of occurence of individual messages at their output. Thus, it is shown in the sequel that the knowledge of only the stationary aspects of the source becomes decisive.

The problem of estimating the efficiency of optimal coding (under any chosen conditions of a priori uncertainty and according to any chosen criterion) and comparison of the results with these estimates was not raised in the present article.

1. COMBINATORIAL CODING FOR A SET OF SIMPLE MESSAGES

1.1 Uniform and "Almost Uniform" Coding

In this section we examine the auxiliary problems of estimating the efficiency of uniform and "almost uniform" coding of a set of messages. We will call "almost uniform coding," the coding in which the lengths of various code combinations differ by not more than unity. Such a consideration is brought about first of all by the fact that in this case the coding and decoding procedures are especially simple. Moreover, the uniform or almost uniform coding is used when nothing is known about the probability of the incidence of individual messages.

Theorem 1. If X is a set of m equally probable messages, then it is possible to set up almost uniform coding so that

$$\overline{n} - H(X) = \overline{n} - \log m \leq \leq \frac{1}{q-1} + \log(q-1) - \log e + \log\log e = f(q) < 1, \quad (1)$$

where q is the base of the code, $\overline{n}$ is the mean number of q symbols used to describe one message of the set X, H(X) is the entropy of the set X determined in the usual manner.*

*In the future the bases of logarithms (contained, for example in (1) and the expression for entropy) are assumed to be equal to the base q of the code.

Proof. Let $q^{s-1} \leqslant m \leqslant q^s$. To describe the messages let us select μ code combinations of lengths s-1. To each of the remaining $q^{s-1} - \mu$ sequences of length s-1 we add one symbol, thus obtaining another $q(q^{s-1} - \mu)$ code combinations. The condition $\mu + q(q^{s-1} - \mu) \geqslant m$ implies $\mu \leqslant (q^s - m/(q - 1)$, and

$$\overline{n} = \mu(s - 1)\frac{1}{m} + (m - \mu)s\frac{1}{m} \leqslant s + \frac{1}{q-1} - \frac{q^s}{(q-1)m}.$$

The difference of the obtained estimate $\overline{n}$ and $H(X) = \log m$ is maximum when $m = q^s/[(q-1) \log e]$ and is equal to $f(q)$. On the other hand it is not difficult to see that $f(q) < 1$. QED

Remark 1. The almost uniform coding examined above is coding according to Khaffmen [Hoffman] [5]. In general it is known that the Khaffmen code is optimal and for it $\overline{n} - H(X) < 1$. Theorem 1 makes it possible in a special case to estimate $\overline{n} - H(X) = \overline{n} - \log m$ more accurately. Thus when $q = \overline{2,10}$, the function $f(q)$ increases from 0.0861 to 0.2689. In other words even when $q = 10$ the value of $f(q)$ is significantly less than one, and when $q = 2$ it is negligibly small.

Theorem 2. If probabilities p_i $(i = \overline{1,M})$ of incidence of messages of the set X satisfy the condition

$$\frac{\varepsilon_1}{m} \leqslant p_i \leqslant \frac{\varepsilon_2}{m}, \quad \varepsilon_1 \leqslant 1 \leqslant \varepsilon_2, \quad i = \overline{1,m}, \tag{2}$$

then in uniform coding

$$\overline{n} - H(X) =]\log m[- H(X) \leq$$

$$\leq \frac{1}{\varepsilon_2 - \varepsilon_1}\left[\varepsilon_2 \log \varepsilon_2 - \varepsilon_1 \log \varepsilon_1 - \varepsilon_1 \varepsilon_2 \log \frac{\varepsilon_2}{\varepsilon_1}\right] + 1 = F(\varepsilon_1, \varepsilon_2) + 1, \quad (3)$$

where $]\alpha[$ denotes the smallest integer, not less than α.

Proof. The maximum of the difference $]\log m[- H(X)$ corresponds to the minimum of H(X). It is easily shown that when condition (2) holds the minimum of H(X) is attained when μ messages have the probability ε_1/m, and $m - \mu$ is the probability of ε_2/m, where $\mu = (\varepsilon_2 - 1)m/(\varepsilon_2 - \varepsilon_1)$ is determined from the normalization condition. Then

$$H(X) \geq -\mu \frac{\varepsilon_1}{m} \log \frac{\varepsilon_1}{m} - (m-\mu)\frac{\varepsilon_2}{m} \log \frac{\varepsilon_2}{m} = \log m - F(\varepsilon_1, \varepsilon_2),$$

and since $]\log m[- \log m < 1$, we obtain (3). QED

In some special cases the expression for $F(\varepsilon_1, \varepsilon_2)$ is significantly simplified:

$$F(\varepsilon_1, \varepsilon_2) \leq F(0, \varepsilon_2) = \log \varepsilon_2, \quad (4)$$

$$F\left(\frac{1}{\Delta}, \Delta\right) = \frac{\Delta - 1}{\Delta + 1} \log \Delta, \qquad \Delta \geq 1,$$

$$F(1-\Delta, 1+\Delta) = \log 2 - \left\{-\frac{1-\Delta}{2} \log \frac{1-\Delta}{2} - \frac{1+\Delta}{2} \log \frac{1+\Delta}{2}\right\}, \quad 0 \leq \Delta \leq 1.$$

To decrease the difference $\overline{n} - H(X)$, it is also possible in this case to use the almost uniform code. Its value may be greater or smaller than $F(\varepsilon_1, \varepsilon_2) + f(q)$ (code combinations of lengths S-1 are "ascribed" at random to messages with probabilities ε_1/m or ε_2/m, correspondingly), but always less than $F(\varepsilon_1, \varepsilon_2) + 1$.

1.2 The Principle of Combinatorial Coding

Let X again be a set of m messages. An arbitrary subset $X_i \subset X$ is called a probability class or simply a class, and the set of nonintersecting probability classes $W = \{X_i,\ i=\overline{1,M}\}$, such that $\bigcup_{i=1}^{M} X_i = X$, is called a set of probability classes or simply a set of classes (for a sequence of messages the latter concept was introduced in [2] under the name of a "complex of messages").

Let m_i be the number of messages in X_i. We number the messages within X_i from 1 to m_i and denote the j^{th} message in X_i by x_{ij}, the probability of its occurence by $p(x_{ij})$, and the probability of occurence of any "representative" of X_i or, in other words, the probability of occurence of X_i in the set of classes W, by $p(X_i)$. Then

$$X = \{x_{ij},\ i=\overline{1,M},\ j=\overline{1,m_i}\},$$

$$X_i = \{x_{ij},\ j=\overline{1,m_i}\},$$

$$\sum_{i=1}^{M}\sum_{j=1}^{m_i} p(x_{ij}) = \sum_{i=1}^{M} p(X_i) = I .$$

Let the q codes $Y_o = \{y(X_i),\ i=\overline{1,M}\}$ and $Y_i = \{y(x_{ij}/X_i),\ j=\overline{1,m_i}\}$ be used to describe the elements of sets W and X_i, respectively ($i=\overline{1,M}$); each of these codes is "divisible"--i.e., it admits unique decoding; $n(X_i)$ and $n(x_{ij}/X_i)$ are the lengths of code combinations $y(X_i)$ and $y(x_{ij}/X_i)$ respectively. To describe a message x_{ij} within the set X (and not only the class X_i) let us construct a code Y so that the following code combinations would correspond to message x_{ij}

$$y(x_{ij}) = y(X_i) \oplus y(x_{ij}/X_i), \quad i=\overline{1,M}, \quad j=\overline{1,m_i}, \tag{5}$$

formulated by the successive recording of $y(X_i)$ and $y(x_{ij}/X_i)$. Its length is

$$n(x_{ij}) = n(X_i) + n(x_{ij}/X_i), \quad i=\overline{1,M}, \quad j=\overline{1,m_i}. \tag{6}$$

The one-to-one correspondence of the elements of sets X and Y is obvious. In addition, there holds

Statement 1. If the codes Y_o and $Y_i (i=\overline{1,M})$, are divisible, then the code Y constructed according to principle (5) is also divisible.

The procedure of successive recording of code combinations of various divisible codes was apparently first used by Fano [5] for coding a Markov source. For unique decoding it is sufficient to know at any given moment to which code the next code combination belongs. In [5] this is determined by the results of decoding of a fixed number (equal to the connectedness of the Markov chain) of preceeding code combinations. In this case in any odd step the code Y_o is used, and in any even step, the code Y_i; moreover, the number of the code i corresponds to the number of the class being determined by the preceeding code combination $y(X_i)$.

The one-to-one correspondence of code combinations $y(x_{ij}) \in Y$ and messages $x_{ij} \in X$ as well as the divisibility of code Y make it possible to use it for describing the set X.

Let us call the set of m_i messages having the probabilities of occurence $p(x_{ij}/X_i) = p(x_{ij})/p(X_i)$ the conditional class X_i^*, and denote by $H(X)$, $H(W)$ and $H(X^*_i)$ the entropies of the sets X, W, and X_i^* determined by the usual method. In addition, let

$$\overline{n(X_i)} = \sum_{i=1}^{M} p(X_i)\, n(X_i), \tag{7}$$

$$\overline{n(x_{ij}/X_i)} = \sum_{j=1}^{m_i} p(x_{ij}/X_i)\, n(x_{ij}/X_i) -$$

be the mean quantity of symbols per message in the coding of the set W with the code Y_o and the conditional class X_i^* with the code Y_i, respectively. Then the following theorem is true.

Theorem 3. In using code Y the mean number of symbols per message of the set X is equal to

$$\overline{n(x_{ij})} = H(X) + \{\overline{n(X_i)} - H(W)\} + \sum_{i=1}^{M} p(X_i)\{n(x_{ij}/X_i) - H(X_i^*)\}. \tag{8}$$

Proof. In view of the definition of $\overline{n(x_{ij})}$, the relation (6) and the obvious equality $p(x_{ij}) = p(X_i)\cdot p(x_{ij}/X_i)$

$$\overline{n(x_{ij})} - H(X) = \sum_{i=1}^{M}\sum_{j=1}^{m_i} p(x_{ij})\, n(x_{ij}) + \sum_{i=1}^{M}\sum_{j=1}^{m_i} p(x_{ij}) \log p(x_{ij}) =$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{m_i} p(x_{ij})\{n(X_i) + n(x_{ij}/X_i) + \log p(X_i) + \log p(x_{ij}/X_i)\} =$$

$$= \sum_{i=1}^{M} p(X_i)\{n(X_i) + \log p(X_i)\} + \sum_{i=1}^{M} p(X_i)\Big\{\sum_{j=1}^{m_i} p(x_{ij}/X_i) \times$$

$$\times \{n(x_{ij}/X_i) + \log p(x_{ij}/X_i)\}\Big\} = \{\overline{n(X_i)} - H(W)\} + \sum_{i=1}^{M} p(X_i) \times$$

$$\times \{\overline{n(x_{ij}/X_i)} - H(X_i^*)\}.$$

which immediately implies (8). QED

Equation (8) determines the "contribution" of non-optimality of each code by the excess of $\overline{n(x_{ij})}$ over H(X). Naturally, to minimize $\overline{n(x_{ij})}$ it is necessary to use codes Y_o and Y_i ($i=\overline{1,M}$) which minimize the mean number of symbols per message of the sets W and X_i^*. But under the conditions of a priori uncertainty two facts are essential: conditional probabilities of occurence of messages are examined within the limits of each class, and in addition, the increase of the difference $\overline{n(x_{ij})}$ - H(X) due to the nonoptimality of the code Y_i is proportional to probability $p(X_i)$. Using the first of these statements, we prove the following theorem.

Theorem 4. If within the limits of each probability class X_i, the probabilities $p(x_{ij})$ are given with accuracy to within an unknown factor, then it is possible to construct a code Y for which

$$\overline{n(x_{ij})} - H(X) \leq \{\overline{n(X_i)} - H(W)\} + I \leq]\log M[- H(W) + I. \quad (9)$$

Proof. Let $p(x_{ij}) = d_i p^*(x_{ij})$, where all $p^*(x_{ij})$ are known, and α_i ($i = \overline{1,M}$) are the unknown factors. But then $p(x_{ij}/X_i)$ which are determined by $p^*(x_{ij})$ are known and are independent of α_i, and hence for the ensemble X_i^* it is possible to select the Khaffmen or the Shannon-Fano code [5], which satisfies the condition

$$\overline{n(x_{ij}/X_i)} \cdot H(X_i^*) < I, \qquad i=\overline{I,M}.$$

Substituting this expression under the summation sign in (8), we obtain the left inequality in (9). Since in selecting $\overline{n(X_i)} =]\log M[$ we obtain $n(X_i) =]\log M[$, the right inequality is also obvious. QED

Remark 2. If each probability class X_i contains equiprobable messages, then by selecting for Y_i $(i=\overline{1,M})$ almost uniform codes, we obtain

$$\overline{n(x_{ij})} - H(X) \leq \{\overline{n(X_i)} - H(W)\} + f(q) \leq \leq \{]\log M[- H(W)\} + f(q) . \quad (10)$$

It is difficult to say for which sets of simple messages the subdivision which satisfies the conditions of Theorem 4 can be made. But the set of sequences of messages at the output of any source of independent messages or Markov source can be divided into classes of equiprobable sequences. This will be utilized later.

1.3 Division Into Classes and Selection of Codes

Let us assume that the set X is divided into classes which satisfy the conditions of Theorem 4. Selecting for Y_i $(i=\overline{1,M})$ appropriate optimal codes, we make sure that inequality (9) is satisfied. If we neglect additions which do not exceed unity the latter means that the nonoptimality of code Y can be related only to the nonoptimality of code Y_o. Therefore, the search for the optimal code Y is reduced to the search for the optimal code Y_o (selection of the lengths of code combinations

$n(X_i)$). Of course it is possible to formulate the problem of selecting a disconnected code Y, which minimizes or bounds $\overline{n(x_{ij})} - H(X)$ for possible distributions of probabilities of individual messages. But when M<<m the search for the code Y_o is significantly simpler (thus in Section 2.7 the selection of an efficient code Y_o for classes of equiprobable successions at the output of a source of independent messages is examined). In addition, if M<<m, then even the uniform code Y_o is efficient.

Even if the messages cannot be divided into classes which satisfy the conditions of Theorem 4, the combinatorial approach often simplifies the search for code Y, which bounds $\overline{n(x_{ij})} - H(X)$. Let us assume that a division into classes has been found in which the utilization of uniform codes Y_i ($i = \overline{1,M}$) $n(x_{ij}/X_i) - H(X_i^*) \leqslant F(\varepsilon_{1i}, \varepsilon_{2i}) + 1 \leqslant \delta_i$, ($\varepsilon_{1i}/m_i$ and ε_{2i}/m_i are the minimal and the maximal possible values of $p(x_{ij}/X_i)$. If, in addition, the selected code Y_o satisfies the condition $\overline{n(X_i)} - H(W) \leqslant \delta_2$, then according to (8), $\overline{n(x_{ij})} - H(X) \leqslant \delta_1 + \delta_2$.

It is obviously appropriate to seek such a division into classes for which the factors $p(X_i)$ $[F(\varepsilon_{1i}, \varepsilon_{2i}) + 1]$ and, accordingly, the "contribution" of each class to the value of $\overline{n(x_{ij})} - H(X)$ are bounded. With this approach the smaller the maximum value of $p(X_i)$ the greater the permissible difference between ε_{1i} and ε_{2i}.

In spite of the fact that the above analysis is of semi-intuitive nature, the possibility of an effective application of the suggested methods is almost obvious in a number of cases

(for example, when the probabilities of occurence of events is governed by the Poisson law with an unknown parameter). However, the general approach to the solution of problems on division into classes and selection of codes Y_o and Y_i is absent.

1.4 The Case in Which Probabilities of Occurence of a Portion of Messages are Known.

If the probabilities of occurence of several messages of the set X are known, then it is appropriate to isolate each of them into a separate class -X_{i_1}, X_{i_2},... and to select $n(X_{i_1}) = -]\log p(x_{i_1})[$, $n(X_{i_2}) = -]\log p(X_{i_2})[$,... (the necessity of numbering the events within these classes is eliminated). All the remaining $y(X_i)$ must be selected by taking into account the already constructed $y(X_{i_1})$, $y(X_{i_2})$,....

If the statistics on the set X are fully prescribed, then with this approach each probability class will contain only one message, all Y_i codes will degenerate, and the coding of probability classes will coincide with the coding of individual messages, thus minimizing the $\overline{n(x_{ij})}$. But then the principle of combinatorial coding itself will also degenerate completely. However even in this case the grouping of events into classes may be appropriate in view of the simplification of the coding and decoding procedures. In any division into classes all $p(x_{ij}/X_i)$ and $p(X_i)$ are known, and Theorem 4 is valid. In addition, it is possible to construct the Khaffmen code for the set of classes W so that $\overline{n(X_i)} - H(W) < 1$, and from (9) it then follows that for a set with known statistics in combinatorial coding one can make sure that $\overline{n(x_{ij})} < H(X) + 2$.

If each X_i contains equiprobable messages, then, according to (10) and the inequality $\overline{n(X_i)} - H(W) < 1$, we obtain $\overline{n(x_{ij})} < H(X) + 1 + f(q)$. Thus, the upper bound of $\overline{n}$ in combinatorial coding is higher than in using the Khaffmen or Shannon-Fano code. It is easy to find examples of sets for which combinatorial coding is worse than optimal coding for any grouping of events into classes. However, the generalization of the principle of combinatorial coding makes it possible to eliminate this difference.

1.5 Generalized Combinatorial Coding

Combinatorial coding can be multilevel. First the messages are divided into classes of the first (lowest) level, then the collection of these classes into classes of the second level, etc., forming a hierarchy of classes. A division in the opposite direction--from higher to lower levels--is also possible. Thus, Khaffmen's coding method for known statistics of messages actually reduces to the formation of q classes of the highest level whose probabilities differ "minimally" from 1/q, and a repetition of this procedure for each class of highest level, etc., until in at least one the classes there would be more than one message. Here each class is coded by one q symbol. Consequently, the generalized combinatorial coding includes also the optimal coding for known statistics of messages.

Under the generalization of combinatorial coding, the relation (8) also changes. Two level combinatorial coding is examined and the corresponding expression is presented in Section II.

II. COMBINATORIAL CODING OF A SEQUENCE OF MESSAGES

2.1 Basic Concepts

Let again $X = \{x_1, x_2, \ldots, x_m\}$ be a set of m messages (which will be referred to as letters from here on), and $\alpha_1, \alpha_2, \ldots, \alpha_j, \ldots$ be the sequence of letters at the output of a discrete source, where each α_j may be any of the m letters of the set X. Let us denote by X^ν the set of all sequences

$$\alpha_j^\nu = \alpha_{j-\nu+1}, \alpha_{j-\nu+2}, \ldots, \alpha_j$$

containing ν letters of the alphabet X, and by $p(\alpha_{j+1}/\alpha_j^\nu)$ and $p(\alpha_{j+1}^{\nu+1})$ the conditional probability of occurence of the letter α_{j+1} for given $\alpha_{j-\nu+1}, \alpha_{j-\nu+2}, \ldots, \alpha_j$ and the combined probability of occurence of the letters $\alpha_{j-\nu+1}, \alpha_{j-\nu+2}, \ldots, \alpha_j, \alpha_{j+1}$. Then

$$p(\alpha_N^N) = p(\alpha_\nu^\nu) \prod_{j=\nu}^{N-1} p(\alpha_{j+1}/\alpha_j^\nu), \qquad N>0, \quad \nu=\overline{1,N}. \tag{11}$$

In addition, let

$$p_\nu(\alpha_N^N) = p(\alpha_\nu^\nu) \prod_{j=\nu}^{N-1} p(\alpha_{j+1}/\alpha_j^\nu), \qquad N>0, \quad \nu=\overline{1,N}, \tag{12}$$

where, as it is easy to see $\sum_{X^N} p_\nu(\alpha_N^N) = 1$, and

$$\begin{aligned} H(X^\nu) &= -\sum_{X^\nu} p(\alpha_\nu^\nu) \log p(\alpha_\nu^\nu), \\ H_\nu(X^N) &= -\sum_{X^N} p_\nu(\alpha_N^N) \log p_\nu(\alpha_N^N), \\ H_\nu(X) &= -\sum_{X^{\nu+1}} p(\alpha_{\nu+1}^{\nu+1}) \log p(\alpha_{\nu+1}/\alpha_\nu^\nu), \end{aligned} \tag{13}$$

where summation is done over all sequences belonging to the corresponding set X^j. It follows from the espressions for p_ν and H_ν that these probabilities and entropies are actually calculated under the assumption that $p(\alpha_N^N) = p_\nu(\alpha_N^N)$ for any N or, in other words, the source "memory" does not exceed ν. From (13) it is easy to obtain that for a stationary source,

$$H_\nu(X^N) = H(X^\nu) + (N-\nu) H_\nu(X) . \qquad (14)$$

2.2 Division Into Classes and the Code Structure

Let us divide the sequence of letters at the source output into blocks of lengths N and examine the set X^N of such blocks. Let

$$\beta^{\nu+1} = \beta_1, \beta_2, \ldots, \beta_\nu, \beta_{\nu+1} \in X^{\nu+1}, \quad \beta^\nu = \beta_1, \beta_2, \ldots, \beta_\nu \in X^\nu ,$$

and for a certain block α_N^N the condition $\alpha_{j+\nu}^{\nu+1} = \beta^{\nu+1}$ is satisfied for $\omega(\beta^{\nu+1})$ values of j $(0 < j \leq N-\nu)$. Then from the definition of a stationary source it follows that the factor $p(\beta_{\nu+1}/\beta^\nu)$ goes into (12) exactly $\omega(\beta^{\nu+1})$ times, and

$$p_\nu(\alpha_N^N) = p(\alpha_\nu^\nu) \prod_{X^{\nu+1}} p(\beta_{\nu+1}/\beta^\nu)^{\omega(\beta^{\nu+1})} \qquad (15)$$

where all $\omega(\beta^{\nu+1})$ correspond to the examined block. Consequently, the probability $p_\nu(\alpha_N^N)$ is completely determined by the assignment of the first ν letters of the block α_N^N and the vector

$$\omega = \{\omega(\beta^{\nu+1}), \beta^{\nu+1} \in X^{\nu+1}\},$$

, and if two blocks have the same first ν letters and the same vectors ω, then the probabilities

$p_\nu(\alpha_N^N)$ of these blocks will be equal regardless of the distribution of probabilities $p(\beta_{\nu+1}/\beta^\nu)$ and $p(\alpha_\nu^\nu)$. Taking into account that various sets of possible vectors ω correspond to various α_ν^ν, let us denote each of them by $\Omega(\alpha_\nu^\nu)$.

Let $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ be the set of blocks α_N^N with prescribed α_ν^ν and ω (all blocks in such a set have the same $p_\nu(\alpha_N^N)$); $\mathcal{Z}(\alpha_\nu^\nu)$ is the set whose elements are all $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ with fixed α_ν^ν and arbitrary $\omega \in \Omega(\alpha_\nu^\nu)$; W is a set, whose elements are all $\mathcal{Z}(\alpha_\nu^\nu)$. Let us consider $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ and $\mathcal{Z}(\alpha_\nu^\nu)$ as probability classes of the first and second level. It is obvious that the probability of occurence $\mathcal{Z}(\alpha_\nu^\nu)$ in the set of classes W is equal to $p(\alpha_\nu^\nu)$. In addition let $p(\omega/\alpha_\nu^\nu)$ and $p(\alpha_N^N/\alpha_\nu^\nu, \omega)$ be the conditional probabilities of occurence of $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ within the class $\mathcal{Z}(\alpha_\nu^\nu)$ and α_N^N within the class $\mathcal{Z}(\alpha_\nu^\nu, \omega)$. Then

$$p(\alpha_N^N) = p(\alpha_\nu^\nu)\, p(\omega/\alpha_\nu^\nu)\, p(\alpha_N^N/\alpha_\nu^\nu, \omega), \tag{16}$$

$$p_\nu(\alpha_N^N) = p(\alpha_\nu^\nu)\, p(\omega/\alpha_\nu^\nu)\, \frac{1}{\tau(\alpha_\nu^\nu, \omega)}, \tag{17}$$

where $\tau(\alpha_\nu^\nu, \omega)$ is the number of blocks in $\mathcal{Z}(\alpha_\nu^\nu, \omega)$.

Since the probability of occurence of $\mathcal{Z}(\alpha_\nu^\nu)$ in the set of classes W is equal to $p(\alpha_\nu^\nu)$ then, according to (13), $H(X^\nu) = H(W)$. As before, let us introduce conditional classes $\mathcal{Z}^*(\alpha_\nu^\nu)$, whose elements $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ have the probabilities of occurence $p(\omega/\alpha_\nu^\nu)$, and entropies $H[\mathcal{Z}^*(\alpha_\nu^\nu)]$.

Let

$$Y(W) = \{ y(\alpha_\nu^\nu),\ \alpha_\nu^\nu \in X^\nu \},\quad Y(\alpha_\nu^\nu) = \{ y(\omega/\alpha_\nu^\nu),\ \omega \in \Omega(\alpha_\nu^\nu) \}$$

и

$$Y(\alpha_\nu^\nu, \omega) = \{ y(\alpha_N^N/\alpha_\nu^\nu, \omega),\ \alpha_N^N \in Z(\alpha_\nu^\nu, \omega) \}$$

be divisible q codes. To describe an arbitrary block α_N^N within the entire set X^N we form code the Y by a successive recording of the corresponding code combinations:

$$y(\alpha_N^N) = y(\alpha_\nu^\nu) \oplus y(\omega/\alpha_\nu^\nu) \oplus y(\alpha_N^N/\alpha_\nu^\nu, \omega), \tag{18}$$

where α_ν^ν and ω correspond to the examined block α_N^N. Repeating arguments presented in Section I, it is possible to show the one-to-one correspondence of blocks $\alpha_N^N \in X^N$ and the set of code combinations $y(\alpha_N^N) \in Y$, and the divisibility of the code Y.

Finally, let us denote by $n(\alpha_\nu^\nu), n(\omega/\alpha_\nu^\nu), n(\alpha_N^N/\alpha_\nu^\nu, \omega)$

and

$$n(\alpha_N^N) = n(\alpha_\nu^\nu) + n(\omega/\alpha_\nu^\nu) + n(\alpha_N^N/\alpha_\nu^\nu, \omega) - \tag{19}$$

the lengths of the code combinations $y(\alpha_\nu^\nu), y(\omega/\alpha_\nu^\nu); y(\alpha_N^N/\alpha_\nu^\nu, \omega)$ and $y(\alpha_N^N)$, respectively, and

$$\begin{aligned}
\overline{n(\alpha_\nu^\nu)} &= \sum p(\alpha_\nu^\nu)\, n(\alpha_\nu^\nu), \\
\overline{n(\omega/\alpha_\nu^\nu)} &= \sum p(\omega/\alpha_\nu^\nu)\, n(\omega/\alpha_\nu^\nu), \\
\overline{n(\alpha_N^N/\alpha_\nu^\nu, \omega)} &= \sum p(\alpha_N^N/\alpha_\nu^\nu, \omega)\, n(\alpha_N^N/\alpha_\nu^\nu, \omega), \\
\overline{n(\alpha_N^N)} &= \sum p(\alpha_N^N)\, n(\alpha_N^N),
\end{aligned} \tag{20}$$

where the summation is over all $\alpha_\nu^\nu \in X^\nu$, $\omega \in \Omega(\alpha_\nu^\nu)$, $\alpha_N^N \in$ $\in Z(\alpha_\nu^\nu, \omega)$ & $\alpha_N^N \in X^N$.

2.3 The General Theorem of Combinatorial Coding for a Sequence of Messages

For the introduced two-level division into classes and the composite code Y it is possible to prove a theorem analogous to Theorem 3.

Theorem 5. In using code Y for describing a sequence of letters of the set X of lengths N, the mean number of symbols $\bar{n} = n(\alpha_N^N)/N$, used to describe one letter, is equal to

$$\bar{n} = H_\nu(X) + \frac{1}{N}\left\{\overline{n(\alpha_\nu^\nu)} - \nu H_\nu(X)\right\} + \frac{1}{N}\sum_{X^\nu} p(\alpha_\nu^\nu)\left\{\overline{n(\omega/\alpha_\nu^\nu)} - H[Z^*(\alpha_\nu^\nu)]\right\} + \frac{1}{N}\sum_{X^\nu} p(\alpha_\nu^\nu)\sum_{\Omega(\alpha_\nu^\nu)} p(\omega/\alpha_\nu^\nu)\left\{\overline{n(\alpha_N^N/\alpha_\nu^\nu,\omega)} - \log z(\alpha_\nu^\nu,\omega)\right\}, \quad (21)$$

Reproduced from best available copy.

Proof. From (16), (19) and (20) it follows that

$$\overline{n(\alpha_N^N)} = \sum_{X^N}\left\{p(\alpha_\nu^\nu)\, p(\omega/\alpha_\nu^\nu)\, p(\alpha_N^N/\alpha_\nu^\nu,\omega)\right\}\cdot\left\{n(\alpha_\nu^\nu) + n(\omega/\alpha_\nu^\nu) + n(\alpha_N^N/\alpha_{\overline{\nu}}^\nu\omega)\right\} =$$

$$= \overline{n(\alpha_\nu^\nu)} + \sum_{X^\nu} p(\alpha_\nu^\nu)\,\overline{n(\omega/\alpha_\nu^\nu)} + \sum_{X^\nu} p(\alpha_\nu^\nu)\sum_{\Omega(\alpha_\nu^\nu)} p(\omega/\alpha_\nu^\nu)\,\overline{n(\alpha_N^N/\alpha_\nu^\nu,\omega)}. \quad (22)$$

On the other hand, according to (17)

$$H_\nu(X^N) = H(W) + \sum_{X^\nu} p(\alpha_\nu^\nu) H[Z^*(\alpha_\nu^\nu)] + \sum_{X^\nu} p(\alpha_\nu^\nu)\sum_{\Omega(\alpha_\nu^\nu)} p(\omega/\alpha_\nu^\nu)\log z(\alpha_\nu^\nu,\omega). \quad (23)$$

Taking into account that $H(W) = H(X^\nu)$, by comparing (14) and (23) we obtain

$$N H_\nu(X) - \nu H_\nu(X) - \sum_{X^\nu} p(\alpha_\nu^\nu) H[Z^*(\alpha_\nu^\nu)] - \sum_{X^\nu} p(\alpha_\nu^\nu)\sum_{\Omega(\alpha_\nu^\nu)} p(\omega/\alpha_\nu^\nu)\log z(\alpha_\nu^\nu,\omega). \quad (24)$$

From (24) and (22) we immediately obtain (21). QED

Corollary. In using uniform codes $Y(\alpha_\nu^\nu, \omega)$, i.e., in selecting $n(\alpha_N^N/\alpha_\nu^\nu, \omega) =]\log \tau(\alpha_\nu^\nu, \omega)[$,

$$\bar{n} - H_\nu(X) < \frac{1}{N}\{\overline{n(\alpha_\nu^\nu)} - \nu H_\nu(X)\} + \frac{1}{N}\sum_{X^\nu} p(\alpha_\nu^\nu)\{\overline{n(\omega/\alpha_\nu^\nu)} - H[\tilde{z}^*(\alpha_\nu^\nu)]\} + \frac{1}{N}. \quad (25)$$

The proof of this corollary is obvious.

Relation (25) makes it possible to obtain estimates from above of the value of $\bar{n}$ for some classes of discrete stationary sources.

2.4 Sources of Independent Messages

Theorem 6. If α_N^N are the blocks of lengths N at the output of a source of independent messages, and the probabilities $p(x_i)$, $i = \overline{1,m}$, are unknown, then in using the combinatorial coding, it is possible to insure that

$$\bar{n} - H(X) < \frac{(m-1)\log N + 2}{N}, \quad N > 3, \quad (26)$$

where H(X) is the entropy of the source of independent messages.

Proof. Let $\nu=0$. For a source of independent messages $H_\nu(X) = H_0(X) = H(X)$. $X^\nu = X^0$ is the empty set, and $\mathcal{Z}(\alpha_0^0) = \{\mathcal{Z}(\omega), \omega \in \Omega\} = \mathcal{Z}$. In other words, classes of the first level are characterized only by the vector ω, classes of the second level degenerate into one class, W and Y(W) are empty sets, and $n(\alpha_0^0) - \nu H_\nu(X) = 0$. In addition,

$$H[\tilde{z}^*(\alpha_0^0)] = -\sum_\Omega p(\omega/\alpha_0^0)\log p(\omega/\alpha_0^0) = -\sum_\Omega p(\omega)\log p(\omega) = H(\Omega),$$

and, since $p(\alpha_0^0) = 1$ (a block always begins with some letter),

$$\sum_{\mathbf{x}} p(\alpha_0^0)\left\{\overline{n(\omega/\alpha_0^0)} - H[\mathfrak{X}^*(\alpha_0^0)]\right\} = \overline{n(\omega)} - H(\Omega).$$

As a result we obtain from (25)

$$\bar{n} - H(X) < \frac{\overline{n(\omega)} - H(\Omega) + 1}{N}. \tag{27}$$

The vector $\omega = \{\omega(\beta),\ \beta \in X\}$ is determined by the set of frequencies of occurence of the letters $x_1, x_2, \ldots, x_m$ in block α_N^N. Since all $\omega(x_i)$ $(i=\overline{1,m})$ are non-negative integers and their sum is equal to N, the total number of different vectors ω is equal to $C_{N+m-1}^{m-1} < N^{m-1}$ $(N > 3)$. By selecting $n(\omega) =]\log C_{N+m-1}^{m-1}[< (m-1)\log N + 1$, we obtain (26) from (27). QED

The estimate (26) was obtained earlier in [3]. In addition, it was conjectured in [2] and proven in [3] that for large N and for $p(x_i)$, $\overline{i=1,m}$, different from zero

$$\bar{n} - H(X) = \frac{(m-1)\log N}{2N} + O\left(\frac{1}{N}\right) \tag{28}$$

(in [4] under the same conditions for m=2 it was shown that it is impossible to insure a smaller value of $\bar{n}$). A small modification of the coding method suggested in [3], made it possible to generalize this result also for the case when some of the $p(x_i)$, $i=\overline{1,m}$ are equal to zero.

2.5 Markov Chains

Theorem 7. If α_N^N are blocks of lengths N at the output

s of a connected Markov chain, and the $p(\alpha_s^s)$ and $p(\alpha_{j+1}^{o}/\alpha_j^s)$ are unknown, then in using the combinatorial coding it is possible to insure that

$$\bar{n} - H(X) < \frac{m^{s+1}\log(N-s)}{N} + \frac{s\log m + 3}{N}, \qquad (29)$$

where H(X) is the entropy of the s-connected Markov chain.

Proof. Let us select $\nu=s$. Then $H_\nu(X) = H_s(X) = H(X)$. The total number of classes of second level is equal to m^s, and for uniform coding $n(\alpha_s^s) < s \log m+1$, and $n(\alpha_s^s) - H(W) < s \log m+1$. The vector ω contains m^{s+1} components $\omega(\beta^{s+1})$. Since all $\omega(\beta^{s+1})$ are non-negative integers and their sum is equal to N-s, the total number of different vectors ω (for all $\beta^{s+1} \in X^{s+1}$) is less than $(N-s)^{m^{s+1}}$, and for uniform coding $n(\omega/\alpha_s^s) < m^{s+1} \log (N-s) + 1$. Then

$$\sum_{X^s} p(\alpha_s^s)\left\{\overline{n(\omega/\alpha_s^s)} - H[\tilde{x}^*(\alpha_s^s)]\right\} \leq \sum_{X^s} p(\alpha_s^s)\,\overline{n(\omega/\alpha_s^s)} < m^{s+1}\log(N-s)+1.$$

Substituting this estimate and the inequality $\overline{n(\alpha_s^s)} - H(W) < s \log m+1$ in (25), we obtain (29). QED

2.6 An Arbitrary Discrete Stationary Source

Theorem 8. For any given arbitrary small positive number ε the sequence of letters, generated by a discrete stationary source with unknown statistics can be coded by means of the combinatorial code Y so that the mean number of symbols per letter of the alphabet X over the set would satisfy the inequality

$$\bar{n} - H_\infty(X) \leq \varepsilon, \qquad (30)$$

where $H_\infty(X)$ is the entropy of a discrete stationary source.

Proof. In accordance with (25) and (29) for an arbitrary stationary source and $\nu=\nu(N)$ it is possible to insure that

$$\bar{n} - H_\infty(X) < \{H_\nu(X) - H_\infty(X)\} + \frac{m^{\nu(N)+1}\log(N-\nu)}{N} + \frac{\nu(N)\log N+3}{N} \quad (31)$$

To prove the theorem it is sufficient to show that it is possible to insure convergence to zero of the right hand side of (31) as $N\to\infty$.

The difference $H_\nu(X)-H_\infty(X) = \varepsilon(\nu) + 0$ as $\nu\to\infty$ [5]. Thus, it is necessary that $\nu(N)\to\infty$ as $N\to\infty$. But simultaneously it is necessary for the second term on the right hand side of (31) to converge to zero, i.e., that

$$\nu(N) = o(\log N - \log\log N) . \quad (32)$$

Actually if ν (N) satisfies (32), then the second, and of course, the third term in (32) converges to zero. Since it is possible to insure that $\nu(N)\to\infty$, while simultaneously satisfying the condition (32), there will always exist an integer N_0 such that when $N > N_0$ each of the three terms on the right hand side of (31) does not exceed $\varepsilon/3$ or, in other words, (30) is satisfied. QED

Remark 3. Condition (32) allows great freedom in the selection of the function $\nu(N)$. But if the dependence $\varepsilon(\nu)$ is known, then for an arbitrary N it is possible to select $\nu(N)$, which minimizes the right hand side of (31). This function $\nu(N)$ insures the most rapid convergence to zero of the right hand side of (31).

Remark 4. If $p(\alpha_N^N) \neq p_\nu(\alpha_N^N)$, then the uniform coding of blocks within $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ is non-optimal and the losses related to this lead to the replacement of $H_\infty(X)$ by $H_\nu(X) \geq H_\infty(X)$ in (21) and (25). But for a source of independent messages and Markov chains $p(\alpha_N^N) = p_\nu(\alpha_N^N)$, if $\nu \geq s$, uniform coding within a class $\mathcal{Z}(\alpha_\nu^\nu, \omega)$ becomes optimal and simultaneously $H_\nu(X)$ becomes $H(X) = H_\infty(X)$.

Remark 5. The estimates obtained above become rougher with the growth of ν. First of all, this is related to the overstating of the number of different vectors $\omega \in \Omega(\alpha_\nu^\nu)$. Not all $m^{\nu+1}$ components of the vector ω are independent, part of them cannot take all the values up to $N - \nu$, and the selection of a fixed α_ν^ν decreases still further the number of possible vectors ω. A certain effect can be obtained also by using nonuniform codes $Y(\alpha_\nu^\nu)$ (the use of the nonuniform code $Y(W)$, is evidently inappropriate).

2.7 An Example of a Nonuniform Code $Y(\alpha_\nu^\nu)$: $\nu=0$, $m=2$

For a source of independent messages the improvement of the estimate can be related only to the utilization of the nonuniform code $Y(\alpha_0^0) = Y = \{y(\omega), \omega \in \Omega\}$. The difference of expressions (26) and (28) is related to the fact that for a uniform code $y(\omega)$ the quantity $\overline{n(\omega)} = \text{const}$, and $H(\Omega)$ depends significantly on the probabilities of occurence of individual letters, and the difference $\overline{n(\omega)} - H(\Omega)$ is the function $p(x_i)$, $i = \overline{1,m}$. Thus, it is necessary to find such a code Y that the difference $\overline{n(\omega)} - H(\Omega)$ would not depend on $p(x_i)$.

Let $m=2$, $p(x_i) = p$, $p(x_2) = 1-p$. Then $\omega = \{\omega_1, \omega_2\} = \{t, N-t\}$, and $\mathcal{Z}(\omega) = \mathcal{Z}(t)$ is the probability class, containing all sequences with $\omega_1 = t$ letters x_1 and $\omega_2 = N-t$ letters x_2.

When $pN = \lambda << N$ the probabilities $p(\omega) = p(t) = C_N^t p^t (1-p)^{N-t}$ are governed by the Poisson distribution, and $H(\Omega) \approx H(\lambda) \approx \frac{1}{2} \log 2\pi e t+1$ [6]. The dispersion of the Poisson distribution is equal to λ, and has the probability approaching unity, $/t-\lambda/ \leqslant \alpha\sqrt{\lambda}$ $(\lambda<1)$. Therefore, by selecting $n(t) = \frac{1}{2} \log 2\pi t + a$ $(t \neq 0$, and a is independent of $t)$, we obtain

$$\frac{1}{2}\log[2\pi(\lambda-\alpha\sqrt{\lambda})] \leqslant \overline{n(t)} - a \leqslant \frac{1}{2}\log[2\pi(\lambda+\alpha\sqrt{\lambda})].$$

If $\lambda>>1$, then $\overline{n(t)} \approx \frac{1}{2} \log 2\pi\lambda + a$, and $\overline{n(t)} - H(\Omega) \approx a$.

To consider the case $t=0$, let us assume $n(t) = \frac{1}{2} \log (2\pi t + 1) + a$. In addition, the above-mentioned arguments are to a greater or smaller degree true for $t \leqslant \frac{N}{2}$ and are not applicable for $t > \frac{N}{2}$. Taking into account the symmetry of $\omega_1 = t$ and $\omega_2 = N-t$, it is natural to select $n(t) = n(N-t)$. In view of these changes and the fact that $n(t)$ is an integer, we select

$$n(t) = \left]a + \frac{1}{2}\log\{1 + 2\pi \min(t, N-t)\right[, \qquad (33)$$

where a is selected in so that the inequality of Kraft is satisfied [5].

This selection of $n(t)$ based on qualitative arguments is nevertheless sufficiently effective.

Theorem 9. If n(t) is defined by the relation (33), and within the limits of each $\mathcal{Z}(t)$ uniform coding is executed, then

$$\bar{n} - H(X) < \frac{\log N}{2N} + \frac{\alpha_0}{N} + \frac{\alpha_1 \log e}{N\sqrt{N}} + \frac{(\pi+2)\log e}{2\pi N^2} \qquad (34)$$

for arbitrary N and $0 \leqslant p \leqslant 1$, where

$$\begin{aligned} \alpha_0 &= 2 + 1{,}5\log 2 - 0{,}5\log\pi, \\ \alpha_1 &= \sqrt{\pi} - \sqrt{1+\pi^{-1}} \end{aligned} \qquad (35)$$

(for q=2 we obtain $\alpha_0 \approx 2{,}680$; $\alpha_1 \log e \approx 0.899$; $\frac{\pi+2}{2\pi}\log e \approx 1.181$).

Proof. By definition

$$\begin{aligned} n(t) - H(\mathcal{Q}) &= \sum_{t=0}^{N} C_N^t p^t (1-p)^{N-t} n(t) + \sum_{t=0}^{N} C_N^t p^t (1-p)^{N-t} \log\left\{C_N^t p^t (1-p)^{N-t}\right\} \leqslant \\ &\leqslant a + I + \sum_{t=1}^{N-1} C_N^t p^t (1-p)^{N-t} \log\left\{q^{n(t)-a-1} C_N^t p^t (1-p)^{N-t}\right\} < \\ &< a + I + \sum_{t=1}^{N-1} C_N^t p^t (1-p)^{N-t} \log\left\{q^{n(t)-a-1} \cdot \frac{1}{\sqrt{2\pi t(1-t/N)}}\right\}, \end{aligned}$$

since, when t=0 and t=N the arguments of the logarithm do not exceed one, and

$$C_N^t p^t (1-p)^{N-t} \leqslant C_N^t \left(\frac{t}{N}\right)^t \left(1-\frac{t}{N}\right)^{N-t} < \frac{1}{\sqrt{2\pi t(1-t/N)}}.$$

From the symmetry of n(t) and $t\left(1-\frac{t}{N}\right)$ with respect to N/2 it follows that

$$\max_{1\le t\le N-1}\left\{q^{n(t)-a-1}\frac{1}{\sqrt{2\pi t(1-t/N)}}\right\}<\max_{1\le t\le \frac{N}{2}}\left\{\sqrt{\frac{2\pi t+1}{2\pi t(1-t/N)}}\right\}<\sqrt{2}\left(1+\frac{1}{2\pi N}\right),$$

$$\log\left\{q^{n(t)-a-1}C_N^t p^t(1-p)^{N-t}\right\}<\log\left\{\sqrt{2}\left(1+\frac{1}{2\pi N}\right)\right\}<\frac{1}{2}\log 2+\frac{\log e}{2\pi N},$$

$$\overline{n(t)}-H(\Omega)<a+1+\sum_{t=1}^{N-1}C_N^t p^t(1-p)^{N-t}\log\left\{q^{n(t)-a-1}C_N^t p^t(1-p)^{N-t}\right\}< \tag{36}$$

$$<a+1+\left(\frac{1}{2}\log 2+\frac{\log e}{2\pi N}\right)\sum_{t=1}^{N-1}C_N^t p^t(1-p)^{N-t}<a+1+\frac{1}{2}\log 2+\frac{\log e}{2\pi N}.$$

It remains to estimate from Kraft's inequality the minimal a. Using the notation of (35) we obtain

$$I=\sum_{t=0}^{N}q^{-n(t)}\le 2\sum_{t=0}^{N/2}q^{-n(t)}\le 2q^{-a}\sum_{t=0}^{N/2}\frac{1}{\sqrt{2\pi t+1}}<$$

$$<2q^{-a}\left[1+\int_{1/2}^{\frac{N+1}{2}}\frac{dx}{\sqrt{2\pi x+1}}\right]=2q^{-a}\frac{1}{\sqrt{\pi}}\left(\sqrt{N+\alpha_2}+\alpha_1\right),$$

where $\alpha_2 = 1 + \pi^{-1}$, or

$$a<\log 2-\frac{1}{2}\log\pi+\frac{1}{2}\log(N+\alpha_2)+\log\left(1+\frac{\alpha_1}{\sqrt{N+\alpha_2}}\right)<$$

$$<\log 2-\frac{1}{2}\log\pi+\frac{1}{2}\log N+\frac{\alpha_2\log e}{2N}+\frac{\alpha_1\log e}{\sqrt{N}}. \tag{37}$$

Substituting a from (37) into (36), and the obtained expression into (27), and taking into account (35), we obtain (34). QED

Remark 6. The last two terms in (21) are of no interest with regard to the asymptotic behavior of $\bar{n}$. But for finite N, it

is essential to obtain a sufficiently accurate estimate, and this is impossible without using additional terms.

References

1. Fitingof, B. M., "Optimal Coding with Unknown and Changing Statistics of Messages" ("Optimal'noe kodirovanie pri neizvestnoj i menyayushchejsya statistike soobshchenij"), Problems of Information Transmission (Problemy peredachi informatsii), 1966, 2, 2, 3-12.

2. Fitingof, B. M., "Compression of Discrete Information" ("Szhatie diskretnoj informatsii"), Problems of Information Transmission (Problemy peredachi informatsii) 3, 3, 28-36.

3. Babkin, V. F., "A Method of General-Purpose Coding of a Source of Independent Messages of Non-Exponential Laboriousness" ("Metod universal'nogo kodirovaniya istochnika nezavisimykh soobshchenij neeksponentsial'noj trudoemkosti"), Problems of Information Transmission (Problemy peredachi informatsii), in print.

4. Krichevskij, P. E., "Relation Between the Excessiveness of Coding and Reliability of Information About the Source" ("Svyaz; mezhdu izbitochnost'yus kodirovaniya i dostovernost'yu svedenij ob istochnike"), Problems of Information Transmission (Problemy peredachi informatsii), 1968, 4, 3, 48-57.

5. Fano, R. M., Information Transmission (Peredacha informatsii) Statistical Theory of Messages (Statisticheskaya teoriya svyazi), Moscow, Mir, 1965.

6. Gordon,J. P., Quantum Effect and Communications Systems. Proc. IRE, 1962, Vol. 50, N 9, 1898-1908.